AD_____
(Leave blank)

Award Number: DAMD170210440

TITLE:

Novel Image Analysis to Link Sub-Nuclear Distribution of Proteins with Cell Phenotype in Mammary Cancer

PRINCIPAL INVESTIGATOR:
(Enter the name and degree of Principal Investigator and any Associates)
David W. Knowles, Ph.D., PI
Sophie A. lelièvre, D.V.M., Ph.D., Colaborator
Fuhui Long, PhD., Post Doctoral Fellow
Jennifer Gong, Under Graduate
Damir Sudar, M.Sc., Colaborator
Mina J. Bissell, Ph.D., Consultant
Sunil Badve, MBBS, MD(Path), FRCPath, Consultant

CONTRACTING ORGANIZATION:
(Enter the Name, City, State and Zip Code of the Contracting Organization)
Lawrence Berkeley National Laboratory, Berkeley, CA 94720

REPORT DATE:
(Enter month and year, i.e., January 2001)
May 2005

TYPE OF REPORT:
(Enter type of report, i.e., annual, midterm, annual summary, final)
Annual Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: (Check one)

☐ Approved for public release; distribution unlimited

☑ Distribution limited to U.S. Government agencies only;
report contains proprietary and unpublished information

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 074-0188

1. Agency Use Only (Leave blank)	2. Report Date	3. Report Type and Period Covered: Annual Report 23rd April 2004 - 22 April 2005
	20 May 2005	

4. Title and Subtitle
Novel Image Analysis to Link Sub-Nuclear Distribution of Proteins with Cell Phenotype in Mammary Cancer

5. Award Number
DAMD170210440

6. Author(s)
David W. Knowles

7. Performing Organization Name (Include Name, City, State, Zip Code and Email for Principal Investigator)
Lawrence Berkeley National Laboratory,
84R0171, 1 Cyclotron Road, Berkeley, CA 94720

E-**Mail: dwknowles@lbl.gov**

8. Performing Organization Report Number (Leave Blank)

9. Sponsoring/Monitoring Agency Name and Address

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. Sponsoring/Monitoring Agency Report Number (Leave Blank)

11. Supplementary Notes (i.e., report contains color photos, report contains appendix in non-print form, etc.)

12a. Distribution/Availability Statement (check one)
☐ Approved for public release; distribution unlimited
☑ Distribution limited to U.S. Government agencies only - report contains proprietary and unpublished information

12b. Distribution Code
(Leave Blank)

13. Abstract *(Maximum 200 Words)* *(abstract should contain no proprietary or confidential information)*

The goal of this project is to develop novel optical imaging / image analysis techniques that will allow automated, quantitative screening to distinguish malignant, pre-malignant, and non-malignant mammary tissue. By quantifying the spatial distribution of these proteins in this relevant culture model system, the work will provide understanding of how such distributions correspond with the phenotype of the cells. In this reporting year we have cultured and studied premalignant S2 cell, co-cultures of premalignant S2 cells and malignant T4 cells and biopsied, paraffin-embedded, thin-sectioned human mammary tissue. We have developed methods to remove the paraffin and fluorescently stain the tissue. We have developed a tool for scoring the accuracy of our automated segmentation analysis which has lead to of improvements to our nuclear segmentation technique. We have developed a foci-counting image analysis technique to measure the nuclear distribution of H4K20m which revealed marked differences in the nuclear distribution of H4K20m in nonneoplastic, premalignant and malignant cells. We are developing statistical, clustering analysis, techniques to group similar nuclei based on their nuclear distribution. The techniques will enable a classification techniques where individual nuclei with in an image will be scored with their probably phenotype on a per cell basis.

14. Subject Terms (keywords previously assigned to proposal abstract or terms which apply to this award)

15. Number of Pages (count all pages including appendices)

16. Price Code (Leave Blank)

17. Security Classification of Report	18. Security Classification of this Page	19. Security Classification of Abstract	20. Limitation of Abstract
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

Novel Image Analysis to Link Sub-Nuclear Distribution of Proteins with Cell Phenotype in Mammary Cancer

Investigators: **David W. Knowles** (Collaborating laboratories: Sudar, Lelièvre, Bissell & Badve)

INTRODUCTION:

The goal of this project is to develop novel optical imaging / image analysis techniques that will allow automated, quantitative screening to distinguish malignant, pre-malignant, and non-malignant mammary tissue. Our hypothesis is that cellular and tissue phenotype is reflected by the organization of components within the nucleus. By quantifying the spatial distribution of these proteins in this relevant culture model system, the work will provide understanding of how such distributions correspond with the phenotype of the cells.

BODY:

The work focuses on the distribution of chromatin-related nuclear proteins in a progression series of cultured HMT-3522 human mammary epithelial cells (HMECs) that mimic early stages of cancer development. Nonneoplastic S1 HMT-3522 cells recapitulate differentiation into phenotypically normal breast glandular structures in 10 day of three-dimensional (3D) culture. 3D culture is preformed by placing cells in contact with an exogenous extracellular matrix enriched in basement membrane components (Matrigel), and supplying cells with essential growth factors and hormones. Glandular structure formation encompasses cell proliferation (until day 6), growth-arrest and deposition of a continuous endogenous basement membrane around the glandular structure. Malignant T4-2 HMT-3522 cells mimic tumor growth, with the formation of disorganized multicellular structures in which cells keep proliferating, when cultured in the same conditions. Nuclear mitotic apparatus, NuMA, protein has been previously found to undergo remarkable changes in its nuclear organization during glandular structure formation (Lelièvre et al., 1998). During the proliferating stage the distribution patterns of NuMA appear homogenous throughout the cell nucleus and was seem similar to what is observed in tumor cells. Whereas, upon differentiation, NuMA is reorganized into foci within the cell nucleus and ring-like patterns at its periphery. Culturing of the cells and fluorescence labeling of the NuMA protein was be carried out, under subcontract, at the Lelièvre laboratory at Purdue University. Image collection, the development of novel image analysis techniques and the image analysis has be performed at our facility at LBNL.

Key outcomes of the first year of work (Months 1 - 12) were the acquisition of images of NuMA organization in cells cultured between 4 and 12 days and the development of a local bright feature (LBF) analysis method. Previously we had only studies proliferating cells at 4 to 5 days of culture, but after 10 days of culture, nonmalignant cells have differentiated into acini and malignant cells have proliferated into large unorganized clusters. Analyzing images from these cultures has strengthened our initial hypothesis about the reorganization of NuMA in nonmalignant cells and this has been key in the further development of our image analysis techniques. The LBF analysis method we developed allows bright foci within a nucleus to be isolated from the diffuse NuMA staining.

In the second year of work (Months 13 - 24) we have maintained our primary focus on quantifying the distribution of NuMA and extended our image analysis capability to allow the distribution of NuMA foci to be measured within an individual nucleus and thus on a per nucleus basis. Briefly, this has been achieved by several key developments. 1) Our nuclear segmentation ability has been extended to allow automated segmentation of individual nuclei from a cluster of cells. 2) Our local bright feature (LBF) analysis has been modified to allow both local bright and local dark features within a nucleus to be isolated for further analysis. 3) Image analysis techniques were developed to allow each nuclear volume to be subdivided into a set of concentric terraces. These refinements have allowed us to calculate the relative density of NuMA foci in each

nucleus, on a per terrace basis and evaluated as an effective radial distribution from the perimeter of the nucleus to its center.

In year 3, (Months 25-36) significant development was done on an automated segmentation analysis which isolated the position and extent of nuclei from 3D images. The local bright feature analysis was refined to extract local dark as well as local bright features from within images of nuclei fluorescently stained for specific nuclear proteins. Compartmentalization of individual nuclei into a set of concentric terraces allowed the quantification of the radial distribution of specific labeled proteins in individual nuclei. Application of these automated imaging techniques to reveal striking differences in the organization of NuMA between proliferating non-malignant cells and proliferating malignant cells. These methods have allowed us to representing the distribution of NuMA bright features associated with different mammary phenotypes as a simple graph, hence enabling an easy interpretation of the spatial distribution of the protein. The performance of the LBF analysis was greatly enhanced by the development of an automated segmentation of the nuclear volume that enabled us to analyze thousands of nuclei in a short period of time. Using this novel image analysis technique we measured the striking reorganization of NuMA during acinar morphogenesis, while no such reorganization occurred during tumor formation. Most importantly, the LBF analysis permitted a clear discrimination between proliferating non-malignant cells and proliferating malignant cells, which was not achieved so far using other evaluation methods.

Significant Progress has been made this year.

1) New Hires:
In May 2004 the PI hired Jennifer Gong an under graduate student in Biomedical Engineering at the University of California. Jennifer was initially hired as a summer student but due to her enthusiasm and productivity the PI has extended her appointment to the present date. Jennifer has become a proficient microscopist and has helped with the image acquisition. Under the direction of the PI, she has also developed a method of reporting on the accuracy of the automated segmentation analysis, which is reported here.

In October 2004 the PI hired Dr. Fuhui Long, who has expertise in vision, statistical analysis and image analysis[1-7].

2) Premalignant S2 Cells and Tissue Co-culture:
In pursuit of our goals to understand the link between nuclear organization and tissue phenotype for a wide range of phenotypes, we have started culturing and studying premalignant S2 cells. When premalignant S2 cells are cultured in 3D they proliferate, growth arrest and form differentiated spheroids of various size. Because premalignant S2 cells differentiate they are nonneoplastic but they are also unlike nonneoplastic S1 cells. Nonneoplastic S1 cells form differentiated spheroids of well defined size which contain a hollow central lumen. A structure known as an acinus. Rather, premalignant S2 cells mimic ductal carcinoma in situ (DCIS) and exhibit a variety of phenotypes which can be clearly established from their size. Small spheroids which are similar in size to nonneoplastic S1 acini (approximately 35 microns in diameter) and rounded, smooth with no lumen. These spheroids are termed S2-Minis. Spheroids with about twice the diameter of the minis with no lumen but still smooth and rounded have been termed S2-Midis. Large spheroids with irregular shape and no lumen are termed S2-Maxis.

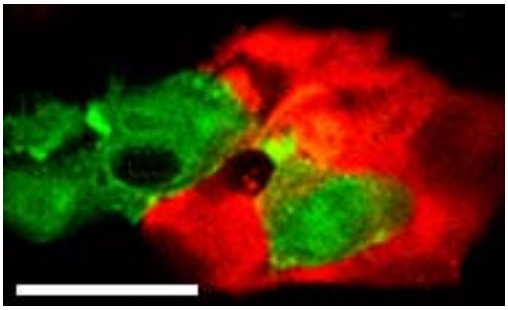


Figure 1. **The use of fluorescent protein expression permits the distinction between non-malignant and malignant cells in co-culture.** In this image malignant dsRed-T4-2 cells (red) have been co-cultured with GFP-nonneoplastic cells expressing GFP (green). Size bar= 5 microns.

Further, as part of the collaboration with the Knowles Laboratory, the Lelièvre Laboratory has recently pioneered methods to co-culture premalignant S2 cells with a small number of malignant T4 cells. The result is that S2 spheroids grow to encapsulate the malignant T4 cells which for independent identification purposes have been transfected with GFP-actin. Figure 1 illustrates the principle that indeed cells from different malignant stages come together to form nodules by showing malignant cells expressing dsRed (red) co-cultured with nonmalignant cells expressing GFP (green).

2) Clinical Samples:

In the past year we have started working with normal, premalignant and malignant human breast tissue with a collaboration with Sunil Badve MBBS, MD(Path), FRCPath, who is Director of Research Immunohistochemistry and Assistant Professor in Pathology at Indiana University. Initially we had worked with freshly frozen tissue because it is most amenable to aqueous-based fluorescent staining procedures we use. However, in preparation for our image-based techniques to be more widely available, we needed to circumvent the problem of working with paraffin-embedded tissues. Paraffin embedding followed by hematoxylin and eosin (H&E) staining are standard histological techniques but paraffin has a large autofluorescent signal and needs to be removed to allow standard aqueous-based techniques of fluorescence immunochemistry.

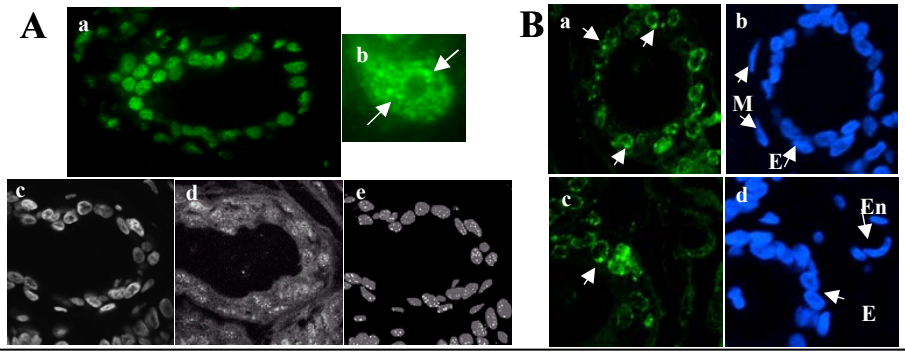


Figure 2. **Fluorescence staining for NuMA and H4K20m on paraffin section of normal breast tissue.** Samples were immunostained for NuMA and H4K20m using streptavidin-biotin-tyramide amplification method. **A**. Low resolution images of NuMA immunostaining (a and d). Arrows indicate the concentration of NuMA bright staining features in mid nucleus (b). DAPI image (c) corresponding to NuMA staining (d). Overlay of the LBF image of NuMA and the segmentation mask (for c and d) is shown in (e). **B.** Low resolution images of H4K20m immunostaining (a and c) and DNA counterstain (b and d). Arrows indicate large H4K20m foci in a and c. E= luminal epithelium; M= myopepithelial cells; En= endothelial cells.

Our current technical analysis shows that a combination of biotin and tyramide gives the brightest specific signal for NuMA and H4K20m on paraffin sections (Figure 2). The interesting aspect of immunostaining combined with DNA counterstaining with DAPI, is that the DAPI staining enables the microscopist to recognize the different regions of the tissue based on the cellular arrangements and also identify landmarks (e.g., blood vessels because of staining- induced red blood cells 'fluorescence') that can be used for comparison between consecutive sections as well as for the precise localization of each nucleus within the sample (Figure

3). This feature will be particularly important when subpopulations of nuclei with different distributions of NuMA and H4K20m will be sought.

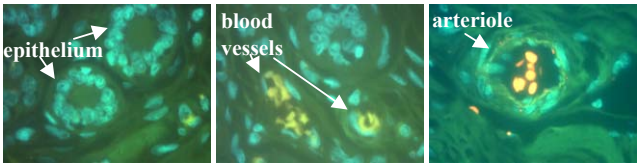


Figure 3. **DAPI staining on paraffin sections of normal breast tissue.** Normal breast tissue was stained for NuMA (FITC-not shown) and DNA (DAPI). DAPI-couterstain highlights nuclei of cell types present in the tissue section, while red blood cells 'fluoresce' in the yellow range (see arrows) and the extracellular matrix appears in pale green. Thus, it is easy to recognize different tissue structures even by fluorescence, such characteristics will be used to position the nuclei analyzed relative to known tissue structures (e.g., blood vessels) and hence map the LBF analysis of each nucleus throughout the tissue section.

1) Analysis of the Accuracy of our Automated Segmentation Technique

In year 2 we developed a novel automated nuclear segmentation technique. Nuclear segmentation is a crucial step in our analysis because it isolates individual nuclei and directs the subsequent nuclear distribution analysis.

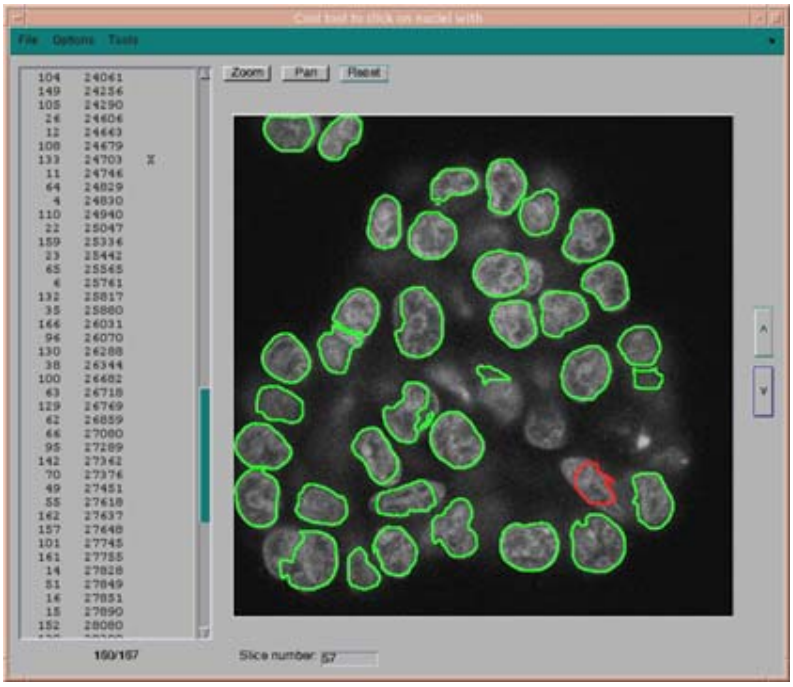


Figure 4. The image shows an interactive tool we have developed for manually scoring the segmentation accuracy. The tool presents a user with the raw nuclear image overlaid with an outline of the corresponding segmentation mask and allows the user to move through the image along the optical direction. The user can choose whether the segmentation overlay is displayed for every nucleus present in the slice or only when the slice corresponding to the center of the nucleus is reached. The progra
m automatically generates and displays a text file which lists the object's number, its volume and its segmentation score, and this file becomes the output of the analysis. The user then simply moves through the images and can repeatedly click on an object to score it. In the figure above, nuclei outlined in green have scored as correctly segmented, and the nucleus outlined in red has been scored as incorrectly segmented. Alternatively, the user can sort the list of objects by number or volume, and by clicking on any object in the list the program automatically updates the image to display the slice corresponding to the center of the nucleus.

One key development this year has been a tool which allows a user to score the accuracy of the segmentation analysis. This tool combines the brilliance of human visual acuity with the speed of the computer and has been essential for revealing weakness in the automated segmentation techniques and has guided a series of developments resulting in a better nuclear segmentation method. This tool allows the results of new segmentation approaches to be quickly compared. Figure 4 shows the user interface of the tools we have developed for manually scoring the segmentation accuracy, which has allowed us to improve our segmentation analysis in significant ways. The tool presents a user with the raw nuclear image overlaid with an outline of the corresponding segmentation mask and allows the user to move through the image along the optical direction.

The user can choose whether the segmentation overlay is displayed for every nucleus present in the slice or only when the slice corresponding to the center of the nucleus is reached. The program automatically generates and displays a text file which lists the object's number, its volume and its segmentation score, and this file becomes the output of the analysis. The user then simply moves through the images and can repeatedly click on an object to score it. In the figure above, nuclei outlined in green have scored as correctly segmented, and the nucleus outlined in red has been scored as incorrectly segmented. Alternatively, the user can sort the list of objects by number or volume, and by clicking on any object in the list the program automatically updates the image to display the slice corresponding to the center of the nucleus.

2) Analysis of the distribution of H4K20 in cultured model tissue.

Encouraged by the success of the nuclear distribution of NuMA, we have developed new techniques to analyze the distribution of another chromatin-related protein, protein histone-4 methylated on lysine-20 (H4K20m). Our observations reveal markedly different staining patterns of H4K20m in nonneoplastic S1 acinar cells and malignant T4 cells (Figure 5).

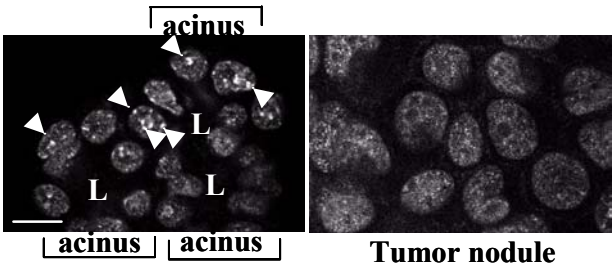


Figure 5. **H4-K20m distribution depends on breast phenotype.** Left image: immunostaining for H4K20m in acini formed in 3D culture (3 acini are shown). Arrows point to large foci of H4K20m. L= lumen. Right image: immunostaining for H4K20m in tumor nodules formed in 3D culture (a portion on a nodule is shown). Size bar= 10 microns.

Notably, acinar cells seemed to possess larger H4K20m staining foci compared to malignant cells. To quantify the nuclear distribution of this protein we have developed a foci-counting algorithm which is based on our LBF technique. In this foci-counting-LBF analysis, a nuclear segmentation mask is first generated from the total-DNA image. Bright foci were then detected and isolated by the LBF analysis from the H4K20m image and scored based on their relative brightness and relative size. Briefly, the score was generated from the LBF image convolved with a narrow Gaussian function.

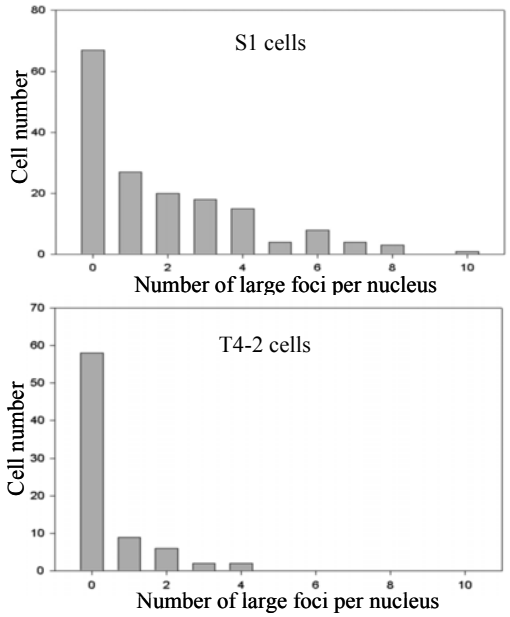


Figure 6. **Histogram of the number of H4K20m foci per nucleus in the S1 acinar cell population and T4 malignant cell population.** S1 and T4 cells were cultured in 3D for 10 days and immunostained for H4K20m. The number of foci per nucleus was determined by our foci-counting-LBF analysis. Histograms of the number of foci per nucleus clearly show that large H4K20m foci are much more abundant in nonmalignant S1 cells than malignant T4 cells.

This manipulation reduced the foci brightness as an inverse function of the foci size. Thus, the local brightness maximum in the blurred-LBF image is a combined measure of the foci brightness and size. The histogram of foci "score" for all nuclei within an image, revealed a large number of small foci but also larger foci at much lower frequency. This result was in agreement with visual analysis of the images. Large foci were defined as having scores larger than two standard deviations away from the histogram mean. To refine the analysis of bright feature distribution, the number of large foci in each nucleus was counted (Figure 6). For a total of 167 nuclei of S1 cells, a total of 311 large foci were counted. Interestingly, 40% of the S1 nuclei were devoid of large foci. In the nuclei that did have large foci there were on average 3.1 large foci per nucleus. For 77 nuclei of T4-2 cells, 35 large foci were counted. 75% of T4 nuclei were devoid of large foci. In the remaining nuclei there were an average of 1.8 large foci per nucleus. The *p* value of the two probability distributions (S1 and T4-2 cells) obtained from the H4K20m analysis was calculated using a Z-test, which compares the significance of the difference between the population means (in this case it is the mean number of large foci per nucleus). For the total number of nuclei the mean number of large foci was 1.86+/-2 for S1 cells and 0.45+/-0.8 for T4 ($Z>7$ and $p<0.001$). Another way to assess if the two distributions are significantly different is to consider the significance of finding a single nucleus with three large foci within the same cell population. The *p*-value for n=3 in the T4 distribution was $p=0.05$ (for a single nucleus with 3 or more foci), whereas the *p*-value for n=3 in the S1 distribution was $p=0.4$ (not significant), as should be the case. The analysis was on per nucleus basis and nuclei with and without large foci could be easily individually isolated.

We then went on to analyze the distribution of H4K20m in premalignant S2 cells. This very recent data clearly shows that the nuclear organization of H4K20m in premalignant S2 cells is markedly different from that in either nonneoplastic S1 cells or malignant T4 cells. Figure 7 below shows a percentage histograms of the number of large H4K20 foci per nucleus for 568 nuclei imaged in multiple S2 spheroids. The data shows that only a small percentage (<10%) of S2 nuclei are without large foci and that on average S2 nuclei have 6.3+-3 large foci. This new data supports our hypothesis by showing that the nuclear distribution of H4-K20m is an excellent marker of nonneoplastic, pre-malignant and malignant phenotypes.

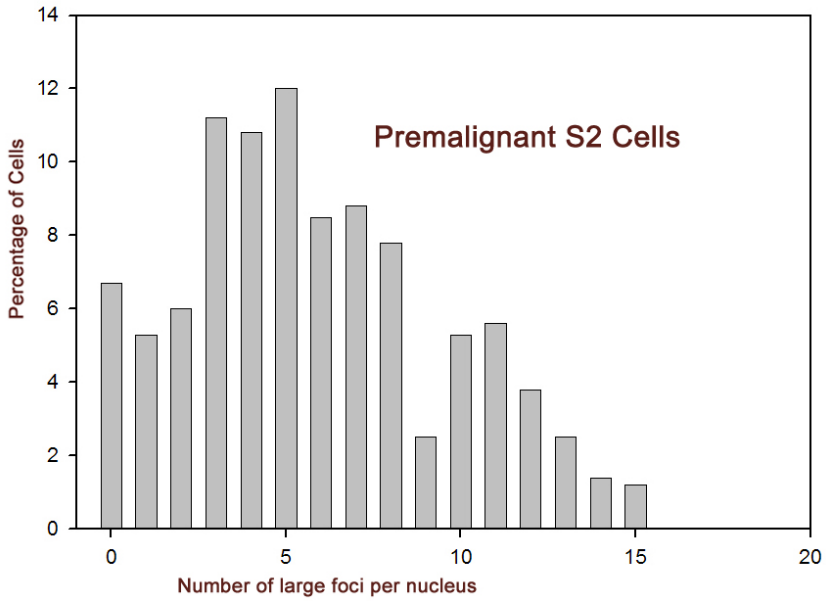


Figure 7. Percentage histogram of the number of large foci per nuclei for premalignant S2 cells show a distribution which is markedly different from that for nonneoplastic S1 cells or malignant T4 cells.

Clustering analysis of the radial distribution of chromatin-associated protein NuMA in nonmalignant S1 and malignant T4 cells.

To this point the results of our LBF analysis have been reported on an average basis. Although this is a powerful way to understand the link between nuclear organization and cell phenotype, it does not directly allow the phenotype of cells to be assigned on a per cell basis. To achieve this we have developed a spectral clustering method to group individual nuclei based on the results of their nuclear organization. To demonstrate this capability, we have grouped thousands of nonneoplastic S1 nuclei, based on the radial-LBF analysis results of their NuMA distribution, and analyzed the statistical link between these clusters and the cell phenotype.

In this method the radial-LBF analysis results of NuMA for each nucleus are represented in a high-dimensional space. The method then computes the Euclidian pair-wise separation for all nuclei in that space. The inverse of these distances is used to calculate the normalized "similarity scores" to describe how similar two nuclei are based on their NuMA distribution. The spectral clustering analysis then uses a hierarchical bipartition scheme to group nuclei into similar clusters based on the "similarity scores". In each iteration, the spectral clustering approach finds an optimal partition that can minimize the similarity between clusters and at the same time maximize the similarity within each cluster. The iteration stops when a metric combining these two criteria converges. Note that this entire procedure is solely based on the NuMA distribution and does not use prior knowledge about the cell phenotype.

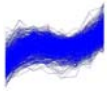
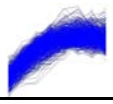
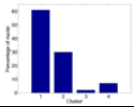
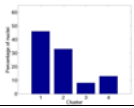
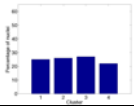
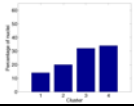
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster Histogram
					
Day 3	61	30	2	7	
Day 5	46	33	8	13	
Day 10	25	26	27	22	
Day 12	14	20	32	34	

Figure 8. Results of the cluster analysis for non-neoplastic S1 cultured for 3, 5, 10 & 12 days as the cells differentiate into acini. Cluster analysis grouped the radial-LBF analysis of NuMA from 2673 nuclei into 4 distinct clusters (Top Row). To understand the link between the clusters and the phenotype, the percentage of nuclei from each day which fall into each cluster is calculated (Rows 2-5) and plotted as a histogram (Right Column). The results show a clear link between clusters and cell phenotype.

Initially we applied this cluster analysis to group 2673 nonneoplastic S1 cells cultured for 3, 5,10 and 12 days into four clusters. The overlaid plot of the NuMA distributions for each cluster is shown by the graphs in the first row of the table (Figure 8). To analyze the statistical link between these clusters and the cell phenotype, we computed the percentage of nuclei falling into each cluster for S1 cells cultured for 3, 5, 10, and 12 days, respectively. The result is shown in the second to fifth row of the table. The results clearly show that: 1) the fraction of nuclei in cluster 1 decreases with the number of days in culture while the fraction of nuclei in cluster 3 and 4 increases; 2) the maximum fraction of nuclei at Day 3 (61%), Day 5 (46%), Day 10 (27%), and Day 12 (34%) fall into cluster 1, 1, 3, 4 respectively; 3) most of the nuclei (over 90%) at Day 3 (corresponding to the proliferation stage) belong to clusters 1 and 2, while most of the nuclei (about 66%) at Day 12 (corresponding to full differentiation into acini) belong to clusters 3 and 4; 4) at Day 12 the fraction of nuclei within each cluster increases from cluster 1 to 4 while at Day 3 the fraction decreases.

By plotting the percentage of nuclei falling into each cluster as a function of the cluster number, we generate a "cluster-histogram" which indicate how nuclei are distributed across clusters for each day, as shown in the last column of the table (Figure 8). The results show a clear statistical correlation between the histogram patterns and the phenotype of cells. In fact, the many clear trends revealed by the clusters generated from spectral clustering method would not have been obtained unless the LBF analysis of NuMA distribution worked correctly and constituted a good marker of cell phenotype.

This analysis was then extended to include all our NuMA distribution data for malignant T4 cells. The cluster analysis for 2673 nonneoplastic S1 cells along with 4418 malignant T4 cells is shown in figure 9. It shows the results of clustering the radial distribution of NuMA for 7091 cells into 20 clusters. The clusters have been arbitrarily ordered. Figure 10 shows the corresponding "cluster histograms" which link the nuclei with in each cluster to their phenotype. As expected the "cluster histograms" show clear difference between nonneoplastic S1 cells cultured for different days but also shows that there is little change in the distribution of NuMA in T4 cells with number of days in culture.

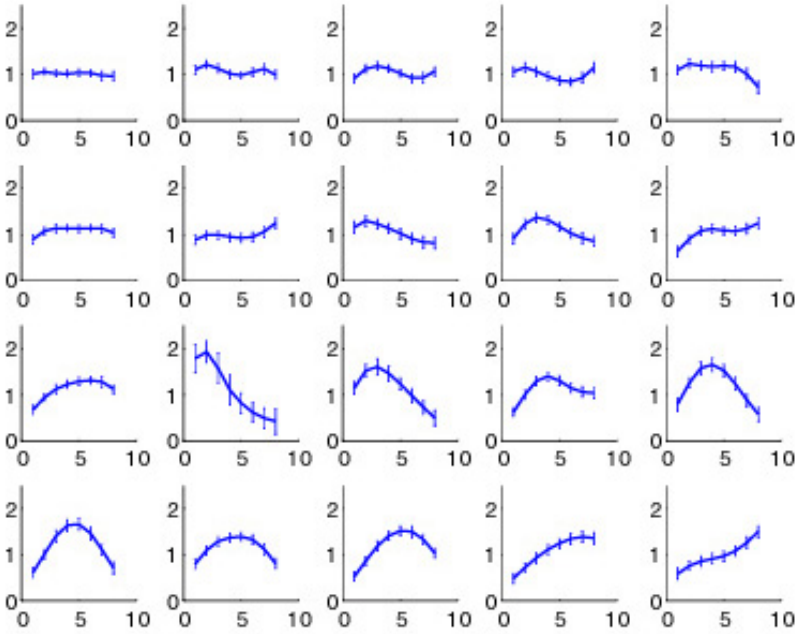


Figure 9: Cluster analysis results of the radial distribution of NuMA in 2673 nonneoplastic S1 cells in combination with 4418 malignant T4 cells.

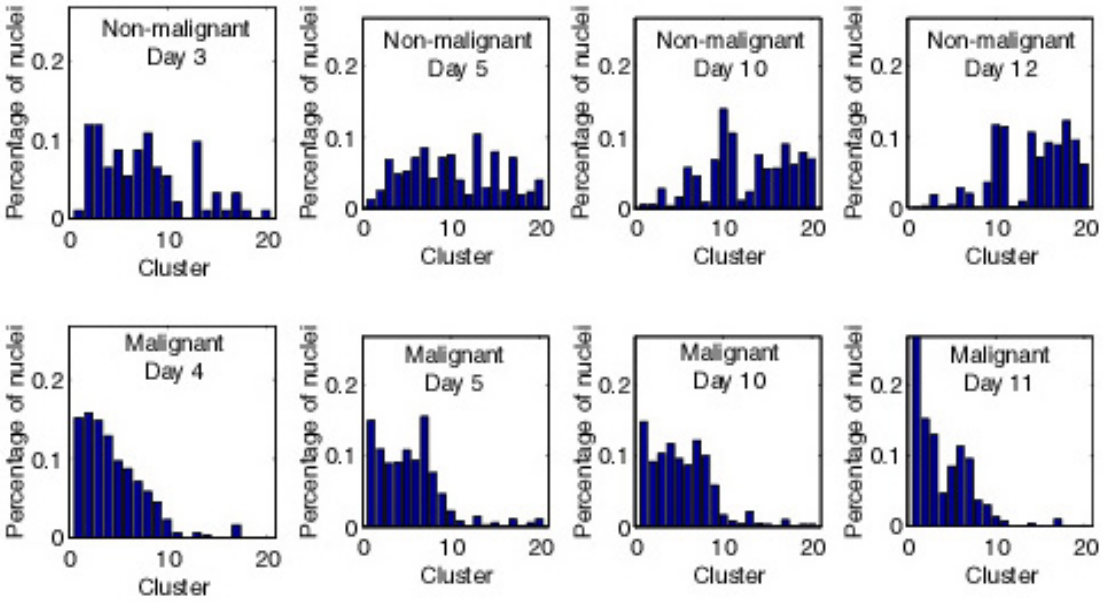


Figure 10: "Cluster histograms" for the 8 possible phenotypes of the 7091 cells being analyzed. There are four possible phenotypes for non-malignant S1 cells (Figure 3 top row), and four possible phenotypes for the malignant cells (Figure 3 bottom row), define by the number of days the cells were in culture.

It is our goal to use this cluster analysis approach to develop a classification method capable of turning high resolution fluorescence images of human mammary epithelial tissue into tissue-maps which report the probable nonneoplastic, premalignant and malignant phenotype at cellular resolution.

KEY RESEARCH ACCOMPLISHMENTS:

* To broaden our knowledge of the link between nuclear distribution and cell phenotype we have cultured and studied premalignant S2 cell and co-cultures of premalignant S2 cells and malignant T4 cells.
* We have established a new collaboration with Sunil Badve MBBS, MD(Path), FRCPath, who is Director of Research Immunohistochemistry and Assistant Professor in Pathology at Indiana University. Dr. Badve has been supplying us with biopsied, paraffin-embedded, thin-sectioned human mammary tissue. We have developed methods to remove the paraffin and fluorescently stain the tissue.
* We have developed a tool for scoring the accuracy of our automated segmentation analysis. The tools has allowed a series of improvements to our nuclear segmentation technique.
* We have developed a foci-counting image analysis technique to measure the nuclear distribution of H4K20m. The results show striking differences in the nuclear distribution of H4K20m in nonneoplastic, premalignant and malignant cells.
* We are developing statistical, clustering analysis, techniques to group similar nuclei based on their nuclear distribution. The techniques will enable a classification techniques where individual nuclei with in an image will be scored with their probably phenotype on a per cell basis. Results from the cluster analysis are reported for the distribution of NuMA in nonneoplastic S1 cells and malignant T4 cells.

REPORTABLE OUTCOMES:

Drs. Knowles and Lelièvre have submitted a manuscript[9].
Drs. Knowles and Long have a paper in preparation[10].
Drs. Knowles and Long have submitted 4 new research proposals since January 2005, based on this work.
Drs. Knowles and Lelièvre have submitted 2 research proposals in the last year, based on this work.

CONCLUSIONS:

The goal of this project is to develop novel optical imaging / image analysis techniques that will allow automated, quantitative screening to distinguish malignant, pre-malignant, and non-malignant mammary tissue. Our hypothesis is that cellular and tissue phenotype is reflected by the organization of components within the nucleus. By quantifying the spatial distribution of these proteins in this relevant culture model system, the work will provide understanding of how such distributions correspond with the phenotype of the cells. The work focuses on the distribution structural nuclear proteins in a progression series of cultured HMT-3522 human mammary epithelial cells (HMECs) that mimic early stages of cancer development.

To broaden our knowledge of the link between nuclear distribution and cell phenotype we have cultured and studied premalignant S2 cell and co-cultures of premalignant S2 cells and malignant T4 cells. We have established a new collaboration with Sunil Badve, Assistant Professor in Pathology at Indiana University, who has been supplying us with biopsied, paraffin-embedded, thin-sectioned human mammary tissue. We have developed methods to remove the paraffin and fluorescently stain the tissue. We have developed a tool for scoring the accuracy of our automated segmentation analysis. The tools has allowed a series of improvements to our nuclear segmentation technique. We have developed a foci-counting image analysis technique to measure the nuclear distribution of H4K20m. The results show striking differences in the nuclear distribution of H4K20m in nonneoplastic, premalignant and malignant cells. We are developing statistical, clustering analysis, techniques to group similar nuclei based on their nuclear distribution. The techniques will enable a classification techniques where individual nuclei with in an image will be scored with their probably phenotype on a per cell basis. Results from the cluster analysis are reported for the distribution of NuMA in nonneoplastic S1 cells and malignant T4 cells.

The developments of the last year represent clear steps forwards towards our goal of using nuclear organization as an epigenetic marker of the phenotype of human mammary epithelial cells. Our current work shows that at least two chromatin-related proteins, NuMA and H4K20m, are particularly sensitive to the transition process which takes normal epithelia into malignancy. Our immediate goal are to validate these findings in human biospied tissue and then to create an image-based screen to identify the phenotypes of cells, on a cell by cell bases, within images human mammary tissue. We believe that such image analysis capability will have an enormous impact in aiding the work of both cancer biologists and oncologists who study and classify cells by epigenetic means.

REFERENCES:

1. Long, F, and Peng, H. (2005) Acquiring and calibrating high dynamic range image database of natural scenes, *Proceedings of IS&T/SPIE International Symposium on Electronic Imaging 2005*, Vol. 5667, 155-161.

2. Peng, H*. Long, F.*, and Ding, C. (2005) Feature selection based on mutual information: criteria of maximal dependency, maximal relevance, and minimal redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.8, 1-13 (*Drs. Peng and Long contributed equally to this work).

3. Long, F., Peng, H., and Feng, D. (2004) Images categorization based on clustering spatial frequency map. *Proceedings of the International Society of Optical Engineering*, 2004, Vol.5307, 212-222.

4. Peng, H., He, X., and Long, F. (2004) Automatic content extraction of filled form images based on clustering component block projection vector. *Proceedings of the International Society of Optical Engineering*, 2004, Vol.5296, 204-212.

5. Peng, H., Long, F., and Chi, Z., (2003) Document image recognition based on template matching of component block projections, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 25(9), 1188-1192.

6. Long, F., and Purves, D. (2003) Scene statistics as a universal basis for color context effects. *Proceedings of National Academy of Sciences of the United States*, 100(25), 15190-15193.

7. Long, F., and Purves, D. (2003) Evidence that color contrast effects have a probabilistic foundation. *Journal of Vision, 3(9). 314.*

8. Long, F., and Purves, D. (2002) A probabilistic explanation of simultaneous brightness contrast, *Journal of Vision, 2(7). 366.*

9. D.W. Knowles, D. Sudar, C. Bator-Kelly, M.J. Bissell, and S.A. Lelièvre. "Automated local bright feature image analysis of nuclear protein distribution identifies changes in tissue phenotype". (*Submitted*)

10. Fuhui Long, Hanchuan Peng, Damir Sudar, Sophie Lelièvre, and David Knowles
Predicting the Phenotype of Mamlian Epithelial Cells using Cluster Analysis of Nuclear Protein Distribution
in preparation.

APPENDICES: